

Solutions of the Ginsparg-Wilson relation and improved domain wall fermions

W. Bietenholz

HLRZ c/o Forschungszentrum Jülich, D-52425 Jülich, Germany

Preprint HLRZ 1998-17, March 1998

We discuss a number of lattice fermion actions solving the Ginsparg-Wilson relation. We also consider short ranged approximate solutions. In particular, we are interested in reducing the lattice artifacts, while avoiding (or suppressing) additive mass renormalization. In this context, we also arrive at a formulation of improved domain wall fermions.

I. The remnant chiral symmetry of fermionic fixed point actions

It is a notorious problem to find a lattice regularization for fermions, which preserves chiral symmetry and avoids species doubling. The celebrated Nielsen Ninomiya No-Go theorem excludes the reproduction of the chiral symmetry manifestly in a local lattice action of a single fermion, if some plausible assumptions are respected [1] (here, locality means that the inverse propagator is analytic in momentum space). However, the manifest chiral invariance of the lattice action is not badly needed to its full extent: with respect to important issues, it is sufficient to preserve just a remnant chiral symmetry, circumventing the Nielsen Ninomiya theorem.

One way to tackle the problem uses the technique of block variable renormalization group transformations (RGTs). Under such transformations, the partition functions and all expectation values – hence the physical contents of a theory – remain invariant [2]. Usually one maps the theory from some fine lattice to a coarse one, increasing the lattice spacing by a factor n . An infinite number of iterations at infinite correlation length leads – for suitable RGT parameters – to a fixed point action (FPA) of the considered RGT. If we send the blocking factor $n \rightarrow \infty$, and we refer to coarse lattice units from the beginning, then we map a continuum theory on the lattice, without altering its physical contents. This mapping, which generates the FPA in just one step, involves a functional integral over the continuum fields.

For free fermions, such a blocking from the continuum can be performed

by the RGT

$$e^{-S[\bar{\Psi}, \Psi]} = \int D\bar{\psi} D\psi e^{-s[\bar{\psi}, \psi]} \times \quad (1)$$

$$\exp \left\{ - \sum_{x, y \in \mathbb{Z}^d} \left[\bar{\Psi}_x - \int dr \Pi(x-r) \bar{\psi}(r) \right] \alpha_{x, y}^{-1} \left[\Psi_y - \int dr' \Pi(y-r') \psi(r') \right] \right\},$$

where $\bar{\psi}$, ψ are the continuum fields, $\bar{\Psi}$, Ψ the lattice fields, s is the continuum action, and S the fixed point lattice action. The matrix α and the convolution function Π specify the RGT (the latter is peaked around 0 and normalized as $\int dr \Pi(r) = 1$). For a massless fermion, we obtain in momentum space the fixed point propagator [3, 4]

$$G(p) = \sum_{l \in \mathbb{Z}^d} \frac{\Pi^2(p + 2\pi l)}{i(p_\mu + 2\pi l_\mu) \gamma_\mu} + \alpha(p), \quad p \in B =]-\pi, \pi]^d. \quad (2)$$

It turns out that the FPA characterized by this propagator is free of fermion doubling [5, 4].

In the limit $\alpha = 0$, the blocking of the fermion fields is implemented by δ functions. In this case, and generally for *any RGT with* $\{\alpha, \gamma_5\} = 0$, the lattice action is chirally invariant ($\{, \}$ denotes the anticommutator). There is no contradiction with the Nielsen Ninomiya theorem, because in all these cases the action is *non-local* [5]: $G^{-1}(p)$ has poles at the corners of the Brillouin zone B . This is a non-locality of the type proposed by Rebhi [6]. His fermion proposal was dismissed, however, because it implied zero axial anomaly [7], and – as an extension of the Nielsen Ninomiya theorem – it has been shown that this is the case for a whole class of non-local fermions with poles in $G^{-1}(p)$ [8]. However, a consistent blocking of the fermionic fields, of the free gauge fields, of the interaction term to the first order in the gauge coupling, and of the axial charge and current, *does* reproduce the axial anomaly correctly on the lattice [9].¹

Now we turn to the case where the action (and the axial current) becomes *local*, so we now deal with $\{\alpha, \gamma_5\} \neq 0$. In this case, which is more interesting in view of practical applications, the RGT term is not chirally invariant, and therefore the chiral symmetry in its naive form is not manifest in the fixed point action. Hence the No-Go theorem does not apply, but the chiral symmetry is still represented correctly *in the observables*, due to the very nature of the RGT [9]. This is an elegant way to by-pass the No-Go theorem.

Since integrating out the continuum fields requires a functional integral, the FPA – and more generally, the perfect action – can be made explicit in perturbation theory. Beyond that, also non-perturbative properties, which are known in the continuum, can be reproduced correctly on the lattice, but

¹The reason why the generalized No-Go theorem does not apply is that the consistent perfect axial lattice current is also non-local, whereas it has been assumed to be local in the proof of Ref. [8].

the expressions for the lattice action seem to become somewhat symbolic. However, it is very interesting that the Atiyah-Singer index theorem still holds even for the classically perfect gauge action [10] (used together with a classically perfect topological charge [11]). Still, the action described in this way (in terms of classical inverse blocking by minimization) is somehow implicit, but certainly not symbolic any more. In the Schwinger model, it was indeed possible to confirm numerically the index theorem for a classically perfect action (to a good accuracy) [12].

The fixed point propagator (2) obeys

$$\{G, \gamma_5\} = \{\alpha, \gamma_5\}. \quad (3)$$

In the case of a local – but not chirally symmetric – term α , this can be viewed as a remnant chiral symmetry of the lattice action. This relation has been postulated by Ginsparg and Wilson as the ‘softest way’ to break chiral invariance in the lattice action [5]. In this form, the remnant chiral symmetry is particularly transparent: it is inherent to the propagator up to a *local violation*.² In particular, it was demonstrated in Ref. [5] that the Ginsparg-Wilson relation (GWR), eq. (3) with local anticommutators, is sufficient to obtain the correct triangle anomaly, and also the soft pion theorems were expected to be reproduced correctly. Furthermore, this property is sufficient to obtain the correct anomaly of $Tr(\gamma_5 G^{-1})$ [10, 14]. In addition, it has been shown that eq. (3) implies a continuous symmetry,³ since the action is invariant under the substitution [14]

$$\bar{\Psi} \rightarrow \bar{\Psi}(1 + \epsilon[1 - G^{-1}\alpha]\gamma_5), \quad \Psi \rightarrow (1 + \epsilon\gamma_5[1 - \alpha G^{-1}])\Psi \quad (4)$$

to $O(\epsilon)$, as we see if we write the GWR in the form

$$\{\gamma_5, G^{-1}\} = G^{-1}\{\gamma_5, \alpha\}G^{-1}. \quad (5)$$

All this sheds light on the “miraculous way” the FPA finds to circumvent the No-Go theorem.

Actually, the transformation (4) is a generalization of the case $\alpha = 1/2$ considered in Ref. [14], to any α permissible in the GWR. However, in the following we also put special emphasis on $\alpha = 1/2$. If we perform the blocking (1) with the standard block average scheme ($\Pi(u) = 1$ if $|u_\mu| \leq 1/2$, $\mu = 1 \dots d$, and $\Pi(u) = 0$ otherwise), then the choice $\alpha = 1/2$ (or $\alpha = -1/2$) does in fact optimize the locality of the free FPA, which we

²In contrast, the usual chiral symmetry breaking terms like mass or a Wilson term correspond to a local violation in the *inverse* propagator. There the symmetry breaking in the propagator becomes non-local. The same happens if we block a massive fermion from the continuum (the mass is added in the denominator of the propagator (2) [13]).

³Blocking a SUSY theory from the continuum could reveal an analogous remnant lattice supersymmetry (but the resulting action differs from the one proposed in Ref. [15]).

write in coordinate space as

$$S[\bar{\Psi}, \Psi] = \sum_{x, r \in \mathbb{Z}^d} \bar{\Psi}_x [\rho_\mu(r) \gamma_\mu + \lambda(r)] \Psi_{x+r} . \quad (6)$$

For this choice, the couplings $\rho_\mu(r)$, $\lambda(r)$ are restricted to nearest neighbors in $d = 1$, and in $d > 1$ their exponential decay is extremely fast.⁴ The numeric values of the leading couplings in $d = 4$ are given in Ref. [13].

If we consider the form (6) as a general ansatz for the lattice action, normalized so that $\sum_r r_\mu \rho_\mu(r) = 1/2$ (*no* sum over μ) and $\sum_r \lambda(r) = 0$ (which means that the fermion is massless), then we see that it is hardly possible to find an ultralocal solution of the GWR (by “ultralocal” we mean that the couplings are restricted to a finite range). In general, λ may have a Dirac structure, but in the following we assume its form to be scalar (in Dirac space). Thus the GWR (5) with $\alpha = \text{const.}$ simplifies to

$$\lambda(r) = \alpha \sum_{y \in \mathbb{Z}^d} [\lambda(y) \lambda(r-y) - \rho_\mu(y) \rho_\mu(r-y)] . \quad (7)$$

For an ultralocal action, the right-hand side extends at least to twice the range of the left-hand side (it extends even further if α is momentum dependent) and it is very unlikely that the latter can achieve cancelations everywhere beyond the range of λ , match with the left-hand side at each point inside that range, and obey the correct normalization.⁵ All this tends to strongly over-determine the degrees of freedom in an ultralocal action. This suggests that in the most local exact 2d and 4d solutions, the couplings in $G^{-1}(r)$ decay exponentially in $|r|$.⁶

Considering eq. (3), one is tempted invent new solutions by hand, such as

$$G(p) = \frac{1}{i\bar{p}_\mu \gamma_\mu} + \alpha , \quad G^{-1}(p) = \frac{i\bar{p}_\mu \gamma_\mu + \alpha \bar{p}^2}{1 + \alpha^2 \bar{p}^2} \quad (8)$$

where \bar{p}_μ is some lattice version of p_μ , and α is a scalar. But for instance the obvious example $\bar{p}_\mu = \sin p_\mu$, which corresponds in fact to a local action for any real α , suffers from fermion doubling. Now one has to remove the doublers, maintaining the locality of $\{G, \gamma_5\}$ (and preferably also of G^{-1}), and it is not trivial to find sensible alternatives to the fixed point propagator [16].

⁴For instance, on the 4-space diagonal the decay behavior is $\rho_\mu(n, n, n, n) \propto \exp(-4.94 \cdot n)$, $\lambda(n, n, n, n) \propto \exp(-4.97 \cdot n)$.

⁵Summation over r leads to the requirement $\sum_y \rho_\mu(y) = 0$. This is guaranteed for the sensible assumption that ρ_μ is odd in y_μ , which holds for the FPA in the block average scheme with an even Dirac scalar α . There, ρ_μ is in addition even in all y_ν , $\nu \neq \mu$, and λ is even in all directions.

⁶The simplification à la eq. (7) for $\alpha = \pm 1/2$ is solved in $d = 1$ by the Wilson action with Wilson parameter $r_W = \pm 1$.

r	$\gamma_5\{G^{-1}(r), \gamma_5\}$	$\gamma_5\sum_x G^{-1}(x)\gamma_5 G^{-1}(r-x)$
$d = 4$		
(0000)	3.70544	3.70544
(0001)	-0.12152	-0.11735
(0011)	-0.06007	-0.06014
(0111)	-0.03194	-0.03356
(1111)	-0.01685	-0.01805
(0002)	0	-0.00417
(0012)	0	0.00020
(0112)	0	0.00069
(1112)	0	0.00047
(0022)	0	-0.00034
(0122)	0	-0.00013
(1122)	0	-0.00008
(0222)	0	-0.00006
(1222)	0	-0.00005
(2222)	0	-0.00002
$d = 2$		
(00)	2.97909	2.97909
(01)	-0.48955	-0.48631
(11)	-0.25523	-0.26035
(02)	0	-0.00324
(12)	0	-0.00350
(22)	0	-0.00188

Table 1: The two sides of the Ginsparg-Wilson relation (5) with $\alpha = 1/2$ for the truncated block average FPA (“hypercube fermion”) in $d = 4$ and in $d = 2$. For all vectors r , which do not occur in the table, both quantities vanish. The exact agreement at zero distance is a consequence of the truncation by periodic boundary conditions.

However, if we can not achieve ultralocality, in practice a truncation of the couplings to a short range is needed, which causes a violation of the GWR. Therefore, what one should aim at is a truncation that keeps this violation small, and here the optimization of locality helps. The truncation can be carried out in an elegant way by means of periodic boundary conditions, which keep the normalization exact. This truncation to couplings in a unit hypercube – for the block average FPA with $\alpha = 1/2$ – has been presented in Ref. [17] (“hypercube fermion”), and its spectral and thermodynamic properties are drastically improved compared to the Wilson fermion [17, 18]. Table 1 shows that also the GWR is approximated very well by this truncated perfect fermion.

This provides hope that the desired properties related to the remnant chiral symmetry are realized to a good approximation for the “hypercube fermion” (HF), in particular the absence of additive mass renormalization.

That property is exact for the perfect and for the classically perfect action [19]. While the index theorem can be considered as classical, it is especially remarkable that also a quantum property like the stability of the chiral limit under renormalization is fulfilled by the classically perfect action. Intuitively, it appears that in the chiral limit the classically perfect actions also displays properties of quantum perfection, and it is plausible that the continuous remnant chiral symmetry (4) protects the zero bare mass from renormalization, similar to the remnant chiral symmetry $U(1) \otimes U(1)$ of the staggered fermions. Here we have the additional virtue that the number of flavors is arbitrary (since doubling is avoided).

As a further virtue of the classically perfect action in the chiral limit, there are no exceptional configurations as shown in Ref. [10] (in the sense they are understood there). However, this paper refers to a fixed point action, which is defined implicitly by minimization. For practical applications, we need a very *simple* gauging prescription (involving only a few short lattice paths), and the requirement to preserve the GWR to a good approximation could guide the construction of such a good but simple gauging. We comment on this concept in the Appendix.

II. A new class of solutions of the GWR

A different solution of the GWR arose from the so-called overlap formalism [20]. To turn it the other way round, for any solution of the GWR, the fermion determinant factorizes in a way compatible with a vacuum overlap [21]. Also this consideration focused on $\alpha = 1/2$, and it is interesting that an entirely different approach singles out the same α as the optimally local block average fixed point fermion.

We re-derive the solution of Ref. [20] and reveal its minimal assumption, which allows for a broad generalization. First we define

$$V = 1 - G^{-1} , \quad (9)$$

and eq. (5) with $\alpha = 1/2$ is equivalent to

$$\gamma_5 V \gamma_5 = V^{-1} . \quad (10)$$

For the case of one flavor, the solution of Ref. [20] reads

$$V = X(X^\dagger X)^{-1/2} , \quad X = 1 - D_W , \quad (11)$$

where D_W is the standard Wilson-Dirac operator.

Depending on the gauge background, there is a danger of a zero eigenvalue of X , so that the expression for V is undefined, but following Ref. [20] we assume that this can be ignored statistically. Note that whenever V is well-defined, then it is unitary. Hence all we need for eq. (10) to hold, is the property

$$\gamma_5 D_W \gamma_5 = D_W^\dagger , \quad (12)$$

since this behavior is inherited by X and V (and by G^{-1}). This leaves a lot of freedom for generalizations. The minimal condition is $\gamma_5 X \gamma_5 = X^\dagger$, but we stay with the form

$$X = 1 - D \quad (13)$$

and note that one may insert many lattice Dirac operators for D . If we make also here the ansatz $D(r) = \rho_\mu(r)\gamma_\mu + \lambda(r)$ – with the normalization mentioned before, and with a scalar form of λ – then this property always holds.

Now the question arises, what generalization is useful. Locality is a criterion, but even if D is ultralocal – as in the case $D = D_W$ – then the decay of the couplings in $G^{-1}(r)$ is still exponential (in $d > 1$). For D_W this can be seen from

$$G^{-1}(p) = 1 - \left[1 - i \sin p_\mu \gamma_\mu - \frac{r_W}{2} \hat{p}^2 \right] \times \quad (14)$$

$$\left[1 + (1 - r_W) \hat{p}^2 + \frac{r_W^2 - 1}{4} \sum_\mu \hat{p}_\mu^4 + \frac{r_W^2}{2} \sum_{\mu > \nu} \hat{p}_\mu^2 \hat{p}_\nu^2 \right]^{-1/2},$$

which simplifies for the standard Wilson parameter $r_W = 1$ and for $r_W = -1$ (the value that the domain wall literature deals with). Locality is realized for any choice of r_W .⁷

Returning to our general ansatz for D , we obtain

$$G^{-1}(p) = 1 - [1 - \rho_\mu(p)\gamma_\mu - \lambda(p)][1 - 2\lambda(p) - \rho^2(p) + \lambda^2(p)]^{-1/2}, \quad (15)$$

and together with eq. (7) this leads to an amusing observation: if we insert an operator D , which obeys the GWR itself, then X becomes unitary, and we end up with

$$G^{-1} = D. \quad (16)$$

For example, the inverse fixed point propagator reproduces itself identically, if we insert it for D . If D violates the GWR, then it gets “GWR corrected”. This correction cannot cure all diseases, however: if D is non-local or plagued by doubling, then the same is true for G^{-1} . On the other hand, also massive fermions can be GWR corrected.

As a further criterion, one should aim at small lattice artifacts. This suggests to insert a choice for D , which has good properties in view of approximate rotation invariance etc., because such properties are essentially inherited by G^{-1} , as the expansion (15) shows. Hence, as an experiment we insert the HF for D . Its small GWR violations (see Table 1) are corrected in G^{-1} , but an exponential tail of couplings is added, so that we obtain something similar to the original fixed point propagator. Of course, this G^{-1} is not fully perfect, but let’s assume that for the aspects of interest

⁷For $r_W = 1/d$ the form of $G^{-1}(r)$ simplifies to $\delta_{r,0}$ + even-odd-couplings, but the latter decay rather slowly. I thank M. Peardon for this remark.

	FPA	HF → GWR	$r_W = 0.85$ → GWR	$r_W = 1$ → GWR	$r_W = 1.15$ → GWR	$r_W = -1$ → GWR
$d = 4$						
r_ρ	1.635	1.519	2.688	2.530	2.537	2.424
r_λ	1.187	1.109	1.819	1.708	1.810	2.784
$d = 2$						
r_ρ	1.268	1.198	1.981	1.816	1.844	1.946
r_λ	0.871	0.844	1.286	1.177	1.206	2.330

Table 2: The locality of the vector and the scalar part of the FPA, of the GWR corrected version of the “hypercube fermion”, and of the GWR corrected Wilson fermion with various Wilson parameters r_W . (The GWR corrected actions are not normalized, because otherwise they violate the GWR again.) As a characteristic quantity, we measure $r_\rho^2 = (\sum_x |\rho_\mu(x)|x^2)/(\sum_x |\rho_\mu(x)|)$, and analogously r_λ .

only the GWR matters. Then the action constructed in this way looks useful, because it is more local than the original FPA (and much more local than the actions obtained from $D = D_W$), see Table 2. Requiring only the GWR means to relax the condition of perfection a little, and we make use of that by further improving locality. For comparison, the leading couplings for some of the fermion actions discussed here are listed in Table 3.

In order to generalize this class of GWR solutions to any (non-trivial) local Dirac scalar α , still assuming $\gamma_5 D \gamma_5 = D^\dagger$, we write

$$X = 1 - 2\sqrt{\alpha} D \sqrt{\alpha}, \quad G^{-1} = \frac{1}{2\sqrt{\alpha}} \left[1 - X(X^\dagger X)^{-1/2} \right] \frac{1}{\sqrt{\alpha}}. \quad (17)$$

Again, G^{-1} fulfills the GWR, and if also D does so, then we end up with $G^{-1} = D$, since X is unitary. The latter further implies that the spectrum of any solution G^{-1} , at a specific momentum $p \in B$, is situated on a circle in the complex plane,

$$\sigma(G^{-1}(p)) \subset \frac{1}{2\alpha(p)}(1 + e^{i\varphi}), \quad \varphi \in [0, 2\pi[. \quad (18)$$

The entire spectrum of G^{-1} is located between the circles given by the minimum and the maximum of $\alpha(p)$.⁸ For our preferred choice $\alpha = 1/2$, the entire spectrum is located on one unit circle. In Figs. 1 and 2 we illustrate that this property is still approximated very well for the FPA after truncation. We compare the HF and its GWR corrected version, which is truncated again in coordinate space, given in the second and in the last

⁸The same geometric structure was obtained in Ref. [10] for the FPA with constant α , but iteration of overlapping blocking functions Π .

	FPA	HF	HF → GWR	$r_W = 1$ → GWR	$r_W = -1$ → GWR	HF → GWR trunc. again
$d = 4$						
$\rho_1(1000)$	0.1400	0.13685	0.1363	0.1668	0.1023	0.13633
$\rho_1(1100)$	0.0314	0.03208	0.0318	0.0239	0.0105	0.03185
$\rho_1(1110)$	0.0096	0.01106	0.0110	0.0050	0.0021	0.01108
$\rho_1(1111)$	0.0033	0.00475	0.0048	0.0007	0.0006	0.00488
$\rho_1(2000)$	0.0040	0	0.0002	0.0238	0.0107	0
$\rho_1(1200)$	0.0015	0	-0.0001	0.0053	0.0017	0
$\lambda(0000)$	1.8530	1.85272	1.8534	1.8334	0.0471	1.85313
$\lambda(1000)$	-0.0616	-0.06076	-0.0626	-0.0560	0.0093	-0.06289
$\lambda(1100)$	-0.0290	-0.03004	-0.0300	-0.0241	0.0027	-0.03024
$\lambda(1110)$	-0.0146	-0.01597	-0.0152	-0.0115	0.0010	-0.01545
$\lambda(1111)$	-0.0077	-0.00843	-0.0079	-0.0057	0.0004	-0.00813
$\lambda(2000)$	0.0010	0	0.0019	-0.0002	-0.0038	0
$\lambda(1200)$	-0.0007	0	-0.0002	-0.0029	-0.0005	0
$d = 2$						
$\rho_1(10)$	0.3154	0.30939	0.3089	0.3330	0.1680	0.30907
$\rho_1(11)$	0.0870	0.09531	0.0953	0.0580	0.0273	0.09540
$\rho_1(20)$	0.0076	0	0.0000	0.0459	0.0288	0
$\rho_1(12)$	0.0048	0	0.0002	0.0166	0.00068	0
$\lambda(00)$	1.4903	1.48954	1.4899	1.4912	0.0651	1.48955
$\lambda(10)$	-0.2441	-0.24477	-0.2464	-0.2335	0.0173	-0.24683
$\lambda(11)$	-0.1246	-0.12761	-0.1252	-0.1167	0.0061	-0.12556
$\lambda(20)$	-0.0009	0	0.0015	-0.0128	-0.0103	0
$\lambda(12)$	-0.0032	0	-0.0017	-0.0108	-0.0028	0

Table 3: The leading couplings of some actions discussed in the text: the FPA, the “hypercube fermion”, the “GWR corrected hypercube fermion” and the “GWR corrected Wilson fermion” with $r_W = 1$ (standard) and $r_W = -1$ (standard for domain wall fermions). All these actions have the symmetries described in footnote 5. The similarity in the first three cases is not surprising, but it is amazing that also the fourth example looks quite similar. For the Wilson fermion we could also vary α according to eq. (17). If we truncate the “GWR corrected hypercube fermion” again periodically, then we reproduce the “hypercube fermion” identically, hence we truncate in coordinate space this time and then correct the normalization by hand; this causes less alteration and yields the last column. (An iteration of GWR correction and truncation does hardly modify the couplings any further.)

column of Table 3. Generally, the eigenvalues are given in $d = 2$ by $\varepsilon_{1,2}(p) = \lambda(p) \pm i\sqrt{\rho^2(p)}$, and in $d = 4$ by the roots of

$$\begin{aligned} &\varepsilon^4(p) - 4\lambda(p)\varepsilon^3(p) + [6\lambda^2(p) + 2\rho^2(p)]\varepsilon^2(p) \\ &- 4\lambda(p)[\lambda^2(p) + \rho^2(p)]\varepsilon(p) + [\lambda^2(p) + \rho^2(p)]^2. \end{aligned}$$

While the progress due to GWR correction is evident in $d = 2$ (Fig. 1), the difference is not easily visible in $d = 4$ (Fig. 2). However, if we measure the “mean deviation” from the unit circle by

$$\delta^2 = \frac{1}{(2\pi)^d} \int_B dp \left[\frac{1}{2^{d/2}} \sum_{i=1}^{2^{d/2}} |\varepsilon_i(p) - 1|^2 \right], \quad (19)$$

then we observe that δ decreases from $1.09 \cdot 10^{-2}$ for the HF to $0.75 \cdot 10^{-2}$ for its GWR corrected and truncated modification. This shows that the latter solves the GWR to an even better approximation.

For the free fermion, $p = 0$ always yields $\varepsilon_i = 0$, $i = 1 \dots 2^{d/2}$, and the whole sector $\varepsilon \approx 0$ is hardly affected by the truncation (see Figs. 1 and 2). If we add a gauge interaction, then this sector is crucial for the index theorem and for the absence of “exceptional configurations”, hence its good quality is very important. Truncation effects are rather visible at the opposite sector $\varepsilon \approx 2$, which corresponds to large momenta.

In a direct application of HFs, it turned out that even using a strongly simplified gauging (instead of the consistently (classically) perfect incorporation of the gauge field), for instance the meson dispersion is drastically improved [17, 22]. However, this application suffers from a strong mass renormalization, which is a problem with respect to practical issues. In particular, for the minimal gauging by hand along shortest lattice paths, the “pion” mass is renormalized from 0.0 to 3.0 at $\beta = 5$ [17]. Thus one can hardly trust the couplings in general, and in addition tuning to the critical bare mass is required. Its sign is opposite to the sign of α , which means that we are led to a region, which is unfavorable for locality. Since the truncation effects in the free fermionic FPA are small, we expect that this renormalization could be almost avoided by using a quasi (classically) perfect gauge interaction, which is, however, very difficult to realize in $d = 4$. If we gauge by hand, trying to suppress the additive mass renormalization (as in Ref. [23]), then one should use the most local free fermion couplings, and here the “GWR corrected hypercube fermion” is in business. For the idea to use the GWR as a guide-line to construct a better gauging by hand, we refer again to the appendix.

III. Improved domain wall fermions

So far we have considered the construction of a potentially interesting GWR corrected propagator for the direct use. However, the way people

think about simulating overlap fermions uses directly D in the above language. In that context, the quality of the locality of G^{-1} is not directly relevant, hence one could choose $D = D_W$ and then gauging is not problematic either. On the other hand, one runs into some trouble with ‘exceptional configurations’ (now in the sense that X has a zero eigenvalue).⁹ More importantly, an extra dimension has to be introduced, which can be interpreted as a fifth direction with L_s sites, or simply as L_s flavors. The meaning of the square root, that the fermion determinant is divided by, is to subtract the contributions of heavy modes, generated by the extra dimension. In practice one has to work with auxiliary boson fields [24].

For practical purposes, the domain wall fermion formulation by Y. Shamir seems most suitable [25], hence we want to discuss possible improvements in that framework. In terms of vector-like gauge theories, first simulations have been performed for the Schwinger model [26, 27] and for QCD [28]. Again, an extra dimension is required, and the limit $L_s \rightarrow \infty$ has attractive features: absence of additive mass renormalization (supported by a one loop calculation [29]), for the measurement of matrix elements no fine tuning is needed either [28], and there are only $O(a^2)$ artifacts¹⁰. In practice L_s must be finite, so that these properties are not exact anymore, but one expects for instance the $O(a)$ artifacts to be suppressed exponentially with L_s . Also the chiral limit appears to be stable to a decent approximation down to $L_s \simeq 10$ [28]. Questions of the L_s truncation have been discussed on the theoretical level in Refs. [25, 30]. We note that Shamir’s formulation allows to work with $L_s/2$ only, in contrast to Kaplan’s original proposal [31].

The action, which has been used so far, has the following form [32]: in Euclidean space – at a fixed extra coordinate s – it is the usual Wilson action with gauge fields on the links, with $r_W = -1$ and a mass M . In the extra dimension, it is the free massless Wilson action with $r_{W,s} = -1$ (the kinetic term with γ_5), and with one additional term

$$D_{s,s'}^{add} = \frac{\sigma}{2} [(1 - \gamma_5)\delta_{s,0}\delta_{s',L_s-1} + (1 + \gamma_5)\delta_{s',0}\delta_{s,L_s-1}] \quad (20)$$

(we refer to one flavor). The bare quark mass is $m_q = M(M - 2)(1 + \sigma)$, but due to renormalization, there is a tuning problem for M . On the other hand, one expects (and observes reasonably well [28]) $m_\pi^2 \propto (1 - \sigma)$.

Now we distinguish three possible strategies for the improvement of this type of domain wall fermions:

1) The observation that certain artifacts of D_W , for instance in the rotational invariance, essentially persist, motivates the use of the HF – or of

⁹The extent of this danger depends on r_W : for $r_W \leq 1$ the smallest eigenvalue of $X^\dagger X$ in the free case is 1, but for larger r_W it decreases down to a minimum of 0.288 at $r_W = 6.4$, before it rises again to 1.

¹⁰This is related to the chiral symmetry at $L_s \rightarrow \infty$, which rules out all $O(a)$ operators. Again, there is a certain similarity with staggered fermions, but here we keep the whole $SU(N) \otimes SU(N)$ symmetry, hence the correct number of Goldstone bosons is involved.

its GWR corrected and truncated modification – in the 4d (or 2d) Euclidean space. Then for example a simplified gauging would be less problematic, because the additive mass renormalization and other diseases are suppressed by L_s . Still, such a simulation would be tedious, but so is the use of a properly gauged HF. If this works, then one cumulates all sort of advantages: very small lattice artifacts (doubly suppressed from the HF and from L_s), a small additive mass renormalization, an arbitrary number of flavors and the correct number of Goldstone bosons. It remains to be checked if the left- and the right-handed fermions can still be separated, see below.

2) Instead one could improve in the 5-direction. If we assume L_s periodicity, then its inverse propagator in the above Wilson-like form reads

$$D_{W,5}^{-1}(p_5) = i \sin p_5 \gamma_5 + \frac{r_{W,s}}{2} \hat{p}_5^2 + \frac{\sigma}{L_s} \exp(i L_s p_5 \gamma_5) . \quad (21)$$

If we block in the 5 direction only, the corresponding “FPA” is given by ¹¹

$$D_5(p_5) = \sum_{l_5 \in \mathbb{Z}} \frac{\Pi_5^2(p_5 + 2\pi l_5)}{(p_5 + 2\pi l_5)(1 + \sigma)\gamma_5 + \sigma/L_s} + \alpha_5 , \quad (22)$$

where Π_5 is the blocking prescription in the 5-direction. If we choose again the block average, and we optimize α_5 for locality, then we arrive at

$$\begin{aligned} D_5^{-1}(p_5) &= (1 + \sigma) \left(\frac{u}{\hat{u}} \right)^2 \left[i \sin p_5 \gamma_5 \pm \left(\frac{1}{2} \hat{p}_5^2 + \hat{u} \right) \right] \\ u &= \frac{\sigma}{(1 + \sigma)L_s} , \quad \hat{u} = e^u - 1 , \quad p_5 = \frac{2\pi}{L_s} j, \quad j = 0 \dots L_s - 1. \end{aligned} \quad (23)$$

In a sense, there is not much to improve in one dimension, since the Wilson action (with $r_W = \pm 1$) is already perfect; we have just incorporated σ and L_s . However, the s dependent defect, which is needed to separate the chiral fermions, has disappeared. A way to avoid this is the use of *fixed* boundary conditions in the s direction, similar to Ref. [33]. This study is in progress.

3) Finally, one could improve directly the entire 5d action, allowing for diagonal couplings with respect to all directions. Since there are no gauge fields on the 5-links, this is perhaps not as disastrous as it first appears. With respect to gauging, there is no additional problem compared to the case 1), and we hope to suppress all discretization artifacts at once. Here, a conspiracy of M , σ and L_s affects the couplings in all directions.

Also in this case, L_s periodic boundary conditions destroy the defect in the s direction ¹², hence a more sophisticated treatment is needed.

¹¹“FPA” needs inverted commas, because we are at finite L_s , but let’s assume that we start at a huge size $n^k L_s$, then perform k RGTs with block factor n , so we end up in L_s . In the limit $n^k \rightarrow \infty$ we obtain the expression (22).

¹²This happens even if we split α^{-1} into $[\alpha_L^{-1}(1 - \gamma_5) + \alpha_R^{-1}(1 + \gamma_5)]/2$, possibly combined with chirally specific functions Π_L, Π_R .

r	$\rho_1(r)$	$\lambda(r)$
	$d = 4$	$M = 1.7$
(0000)	0	0.95860
(1000)	0.02606	-0.01613
(1100)	0.00468	-0.00476
(1110)	0.00123	-0.00178
(1111)	0.00041	-0.00076
	$d = 2$	$M = 0.9$
(00)	0	1.14008
(10)	0.14008	-0.10247
(11)	0.03278	-0.04382

Table 4: The truncated perfect “hypercube fermion” at the masses, which have been used in domain wall fermion simulations: $M = 1.7$ in $d = 4$, and $M = 0.9$ in $d = 2$.

The defect in the s direction is needed to keep the chiral fermions apart, so we focus on variant **1**), which leaves the couplings between the 4d layers untouched. Since in general a mass parameter M appears in the physical subspace, we insert a massive HF. This is constructed from the perfect action at mass M , again by truncation with periodic boundary conditions [17]. There are two choices for α which restrict the 1d action to nearest neighbors, $\alpha = \alpha_M = (e^M - M - 1)/M^2 > 0$, and $\alpha = -\alpha_M$. In higher dimensions, the first (second) option optimizes the locality for positive (negative) M .¹³ In the following, we choose $\alpha = \alpha_M$. As a sound basis, the hypercubic couplings for typical masses, which have been used in simulations, $M = 1.7$ in $d = 4$ (the approximate (quenched) critical value at $\beta = 6$ [28]) and $M = 0.9$ in $d = 2$ [26, 27], are given in Table 4. Note, however, that the renormalization with an improved action could lead to a different critical value of M .

Finally, we want to show that it is indeed possible in this framework to separate the chiral modes. To be able to do so analytically, we now let $L_s \rightarrow \infty$, and we switch to Kaplan’s version, with an ordinary Wilson action in the s direction plus a mass-like term $m(s) = m \cdot \text{sign}(s)$, $\text{sign}(s) = 1$ for $s > 0$, -1 for $s < 0$, and 0 at $s = 0$. Hence the free fermion action reads

$$\begin{aligned}
S[\bar{\Psi}, \Psi] = & \sum_{x,r,s,s'} \bar{\Psi}_{x,s} \left[[\rho_\mu(r)\gamma_\mu + \lambda(r)]\delta_{s,s'} + \left[\frac{1}{2}(\delta_{s,s'+1} - \delta_{s+1,s'})\gamma_5 + m(s) \right. \right. \\
& \left. \left. + \frac{r_{W,s}}{2}(\delta_{s,s'+1} + \delta_{s+1,s'} - 2\delta_{s,s'})\right]\delta_{r,0} \right] \Psi_{x+r,s'} , \quad (24)
\end{aligned}$$

where ρ_μ, λ refer to the HF at some mass M . We now follow the procedure of Ref. [34] to look for solutions of the type $\Psi_{R,L} = e^{ipx}\Phi_s u_{R,L}$, where

¹³Generally, the changes $M \rightarrow -M$ and $\alpha \rightarrow -\alpha$ imply $\rho_\mu \rightarrow \rho_\mu, \lambda \rightarrow -\lambda$.

$\gamma_5 u_R = u_R$, $\gamma_5 u_L = -u_L$. We require the inverse propagator $G^{-1}(p, s)$ to reduce to a chirally invariant $G^{-1}(p)$,¹⁴ which amounts to the condition

$$\left[\frac{1}{2}(\delta_{s,s'+1} - \delta_{s+1,s'})\gamma_5 + m(s) - \frac{r_{W,s}}{2}(\delta_{s,s'+1} + \delta_{s+1,s'} - 2\delta_{s,s'}) + \lambda(p) \right] \Phi_s u_{R,L} = 0. \quad (25)$$

Following Kaplan [31] we make the ansatz $\Phi_{s+1} = z\Phi_s$ for $s, s+1 \neq 0$. The solutions for z are

$$z_{R,L} = \frac{U \pm \sqrt{U^2 - r_{W,s}^2 + 1}}{r_{W,s} \mp 1}, \quad U = m(s) + \lambda(p) + r_{W,s}. \quad (26)$$

The upper (lower) sign in the denominator refers to z_R (z_L), and in both cases the two signs in the numerator are possible. As an example, in the limit $r_{W,s} \rightarrow 1$ there is only one finite solution: $z_R = 1/U$. Normalizability requires

$$\begin{aligned} |z_R(s > 0)| &= \left| \frac{1}{1 + m + \lambda(p)} \right| < 1 \rightarrow \lambda(p) > -m \text{ or } \lambda(p) < -(2 + m) \\ |z_R(s < 0)| &= \left| \frac{1}{1 - m + \lambda(p)} \right| > 1 \rightarrow (m - 2) < \lambda(p) < m. \end{aligned} \quad (27)$$

The allowed region for $\lambda(p)$ is shown as shaded areas in Fig. 3 on top. Below we show the interval of values that $\lambda(p)$ actually takes, depending on M (for $\alpha = \alpha_M$). The minimum is at $\lambda(p = 0) = M^2/(e^M - 1)$, and we see that there are many possible combinations of m and M which yield a single right-handed fermion. In some cases the solution collapses at some “critical momenta”, in others it extends all over the Brillouin zone. It is particularly favorable to choose a rather large parameter M , so that 4d locality is excellent and – as a related property – $\lambda(p)$ is confined to a narrow interval for all momenta. Then, for instance $m \approx 1 \dots 2$ guarantees the solution to exist at any momentum.

In the limit $r_{W,s} \rightarrow -1$ the situation is analogous, but there one deals with a left-handed fermion.

IV. Conclusions

We have revisited the fixed point action approach to the formulation of lattice chiral fermions, emphasizing in particular the blocking from the continuum and the rôle of the Ginsparg-Wilson relation. We also consider the artifacts in the remnant chiral symmetry in short ranged truncated fixed point actions. Based on a formula used in the overlap formalism, we found

¹⁴Here we mean full chiral invariance, not just the GWR. Hence $G^{-1}(p)$ is not a truncated perfect Dirac operator, but just its vector part $i\rho_\mu(p)\gamma_\mu$, which – in its isolated form – would imply fermionic doubling.

a method how a lattice Dirac operator can be “GWR corrected”. This leads to a large class of new solutions, and it allowed us to further optimized their locality, which pays off in reduced spectral artifacts after truncation. Along these lines, we also arrived at a formulation of improved domain wall fermions. We discussed practical issues in view of their application in simulations of vector-like theories. The improvement could, for instance, reduce the extent L_s in the fifth direction, which is needed for a number of favorable properties. Finally, we showed that the mechanism to separate the chiral fermions still works in our improved formulation.

Acknowledgment I am indebted to S. Chandrasekharan, F. Niedermayer, K. Orginos, M. Peardon and U.-J. Wiese for interesting comments.

Appendix

We know that FPAs obey the GWR, which implies plenty of attractive properties, but we also know that such actions are extremely hard to implement. What we need is a good but very simple gauging of our HF, and here we want to comment briefly on the possibility to use the GWR for the construction of such a gauging.

We denote the lattice Dirac operator by D and refer to $\alpha = 1/2$, so that the GWR can be written as

$$D(r) + \bar{D}(r) = \sum_x \bar{D}(x) D(r-x) , \quad \bar{D} = \gamma_5 D \gamma_5 . \quad (28)$$

Let $D(r)$ represent the HF in the free case, which obeys the GWR to a good accuracy (see Table 1). The task is now to find a gauging which preserves this property in the interacting case. This is rather involved, because now the GWR is supposed to hold approximately for each lattice path. (The situation is even worse in cases where $\alpha \neq const.$).

As a simple example, we consider the gauging by the average of the shortest lattice paths only, and we discuss the case $d = 2$. Regarding condition (7), the case $r = (00)$ remains exact, and for $r = (11)$, (21) and (22) the combination of couplings, which nearly matches in the free case, acts on all paths involved, hence the GWR is still approximated well. For $r = (20)$ that combination of couplings splits into coefficients for the paths of length 2 and 4, but all those coefficients are small, so the situation is not much worse than in the free case. The main problem is the case $r = (10)$, which amounts to

$$-0.490 [link] = -0.608 [link] + 0.061 [sum\ over\ staples] . \quad (29)$$

Here the matching path by path does not work well, and this provides some insight into the limitation of the minimal gauging by shortest lattice paths.

One should now correct for that by inserting fat links, a clover term etc. and tuning its coefficients. This work – and the extension to $d = 4$ – is in preparation.

At this point, we just add a remark on how to simplify this task. If we generally define $X = 1 - D$, $\bar{X} = \gamma_5 X \gamma_5$, then the GWR for $\alpha = 1/2$ reads

$$\bar{X}X = 1 . \quad (30)$$

If D is linear in the γ_μ 's, $D = \rho_\mu \gamma_\mu + \lambda$, then the GWR is identical to the requirement that X be unitary. We insert again the HF for D and introduce a new hypercubic variable $\tilde{\lambda} = 1 - \lambda$, so the GWR takes the form

$$\sum_x [-\rho_\mu(x)\rho_\mu(r-x) + \tilde{\lambda}(x)\tilde{\lambda}(r-x)] = \delta_{r,0} . \quad (31)$$

Again this holds exactly at $r = 0$ for any reasonable gauging, but $r \neq 0$ is complicated. Of course, we cannot avoid the inconvenient quadratic term by these re-definitions, but at least the rest has been optimally simplified.

References

- [1] H. Nielsen and M. Ninomiya, Nucl. Phys. B185 (1981) 20 [Erratum B195 (1982) 541]; B193 (1981) 173.
- [2] K. Wilson and J. Kogut, Phys. Rep. C12 (1974) 75.
- [3] M. Peskin, Cornell Univ. Report CLNS-396 (1978). D. Carpenter, DESY 85/064 (1985).
- [4] U.-J. Wiese, Phys. Lett. B315 (1993) 417.
- [5] P. Ginsparg and K. Wilson, Phys. Rev. D25 (1982) 2649.
- [6] C. Rebbi, Phys. Lett. B186 (1987) 200.
- [7] G. Bodwin and E. Kovacs, Phys. Lett. B193 (1987) 283. M. Campostrini, G. Curci and A. Pelissetto, Phys. Lett. B193 (1987) 279.
- [8] A. Pelissetto, Ann. Phys. 182 (1988) 177.
- [9] W. Bietenholz and U.-J. Wiese, Phys. Lett. B378 (1996) 222; Nucl. Phys. B (Proc. Suppl.) 47 (1996) 575.
- [10] P. Hasenfratz, V. Laliena and F. Niedermayer, hep-lat/9801021.
- [11] M. Blatter, R. Burkhalter, P. Hasenfratz and F. Niedermayer, Phys. Rev. D53 (1996) 923.
- [12] F. Farchioni and V. Laliena, hep-lat/9802009.

- [13] W. Bietenholz and U.-J. Wiese, Nucl. Phys. B464 (1996) 319.
- [14] M. Lüscher, hep-lat/9802011. This paper also shows that, starting from a solution of the GWR, one can build a lattice Yukawa theory, which is chirally invariant in the ordinary sense but non-local, similarly to the FPA at $\alpha = 0$.
- [15] T. Aoyama and Y. Kikukawa, hep-lat/9803016.
- [16] Since the lattice Dirac operator $D_\chi(p) = i\bar{p}_\mu\gamma_\mu$ is chirally invariant, it must violate an assumption of the Nielsen-Ninomiya theorem. Most violations – such as doubling, non-Hermiticity, non-technicality, non lattice translation invariance, or non-locality in the sense that $D_\chi(p)$ performs finite gaps – are still present in $G^{-1}(p)$ and therefore unfavorable. However, we can allow for a non-locality in the sense that $D_\chi(p)$ has divergences, and still obtain a local $G^{-1}(p)$. This is exactly the mechanism to proceed from a δ function FPA to a local FPA.
- [17] W. Bietenholz, R. Brower, S. Chandrasekharan and U.-J. Wiese, Nucl. Phys. B (Proc. Suppl.) 53 (1997) 921.
- [18] W. Bietenholz and U.-J. Wiese, hep-lat/9801022, to appear in Phys. Lett. B.
- [19] P. Hasenfratz, hep-lat/9802007.
- [20] H. Neuberger, hep-lat/9801031.
- [21] R. Narayanan, hep-lat/9802018.
- [22] K. Orginos et al., Nucl. Phys. B (Proc. Suppl.) 63 (1998) 904.
- [23] T. DeGrand for the MILC collaboration, hep-lat/9802012.
- [24] R. Narayanan, H. Neuberger and P. Vranas, Phys. Lett. B353 (1995) 507.
- [25] Y. Shamir, Nucl. Phys. B406 (1993) 90.
- [26] A. Jaster, hep-lat/9605011.
- [27] P. Vranas, Nucl. Phys. B (Proc.Suppl.) 53 (1997) 278.
- [28] T. Blum and A. Soni, Phys. Rev. D56 (1997) 174; Phys. Rev. Lett. 79 (1997) 3595.
- [29] S. Aoki and Y. Taniguchi, hep-lat/9709123.
- [30] H. Neuberger, hep-lat/9710089. Y. Kikukawa, H. Neuberger and A. Yamada, hep-lat/9712022.

- [31] D. Kaplan, Phys. Lett. B288 (1992) 342.
- [32] V. Furman and Y. Shamir, Nucl. Phys. B439 (1995) 54.
- [33] E. Katz and U.-J. Wiese, comp-gas/9709001.
- [34] K. Jansen and M. Schmaltz, Phys. Lett. B296 (1992) 374.

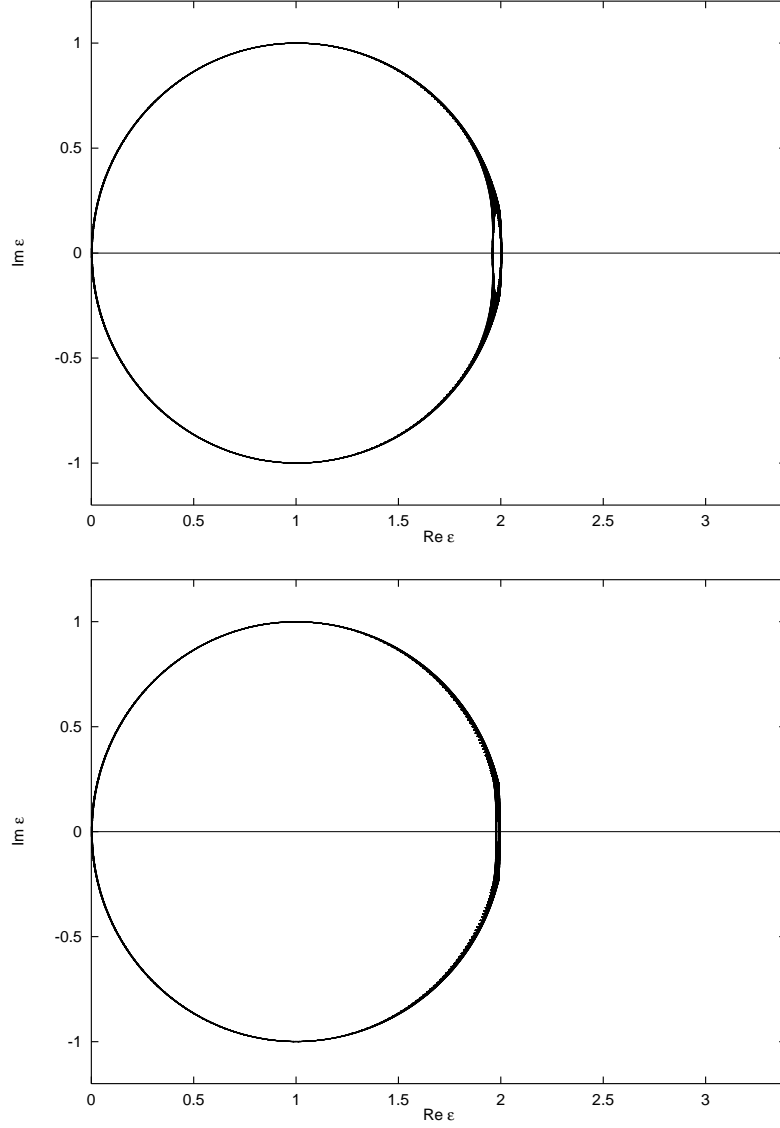


Figure 1: Spectrum of the “hypercube fermion” (on top) and its GWR corrected and truncated modification (below) in $d = 2$. In both cases, the spectra keep close to the unit circle, hence the artifacts due to truncation are small. In the most contaminated region around $\varepsilon \simeq 2$ we observe some progress thanks to GWR correction.

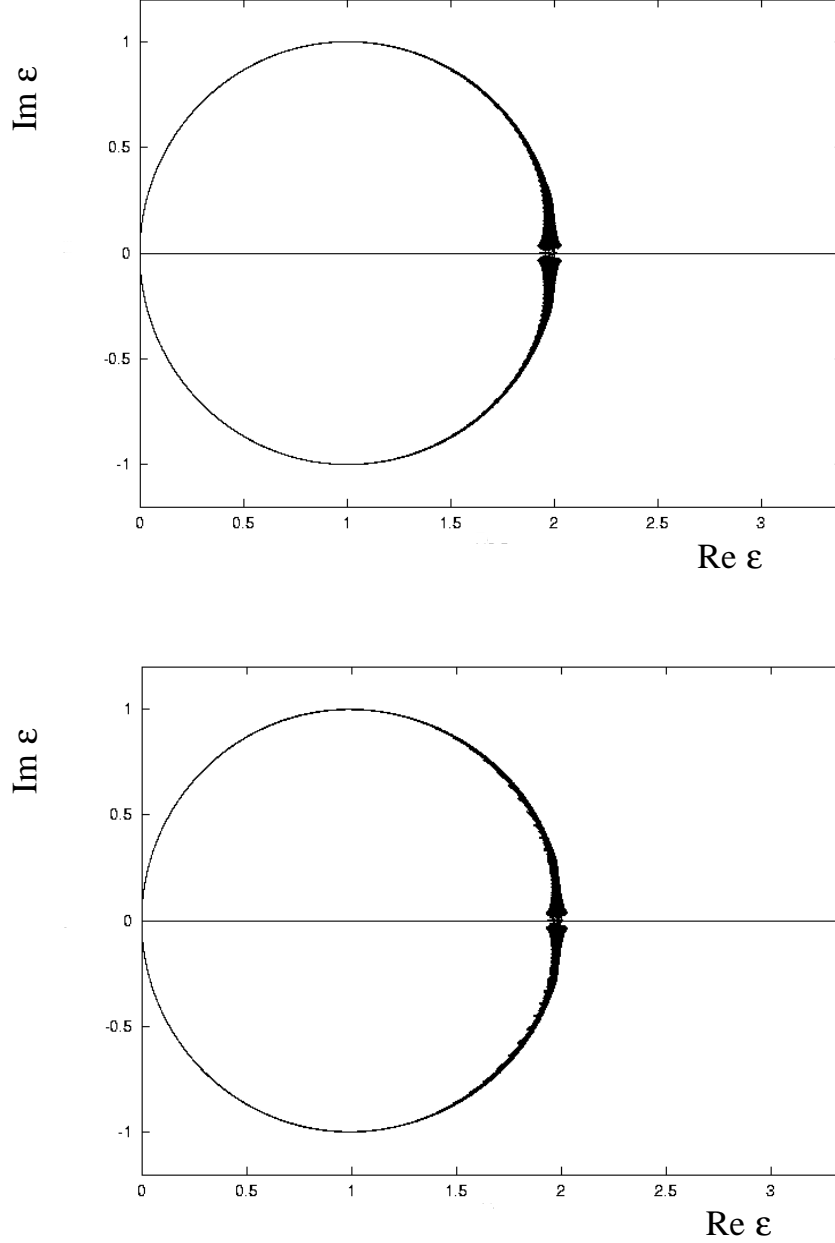


Figure 2: Spectrum of the “hypercube fermion” (on top) and its GWR corrected and truncated modification (below) in $d = 4$. Again, the spectra keep close to the unit circle. From the shape the progress due to GWR correction is not easily visible here, but if we measure the mean deviation from the unit circle, which takes into account the eigenvalue *density*, then we observe an improvement by 31 percent.

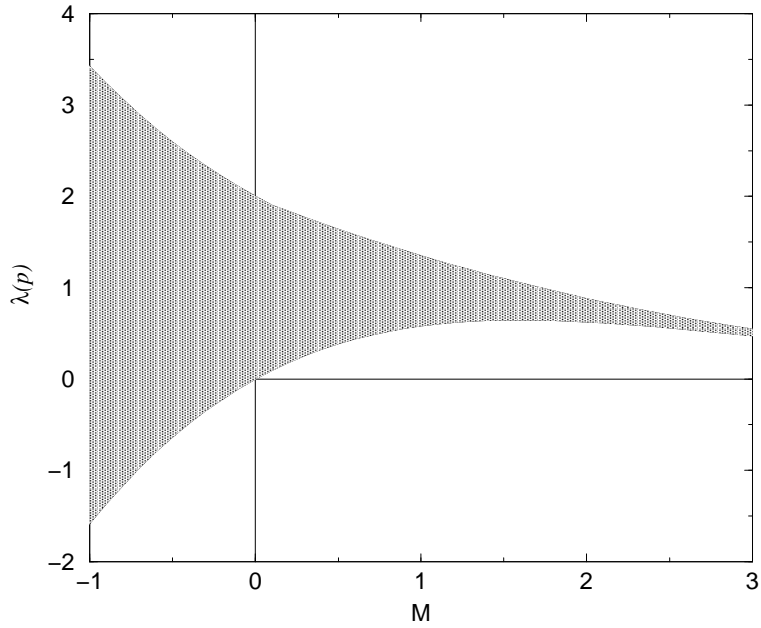
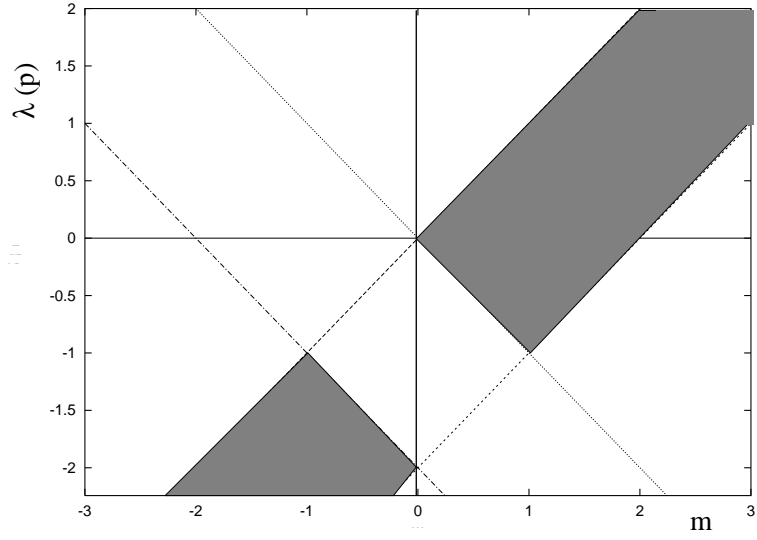


Figure 3: On top, the shaded regions are the allowed areas for $\lambda(p)$ providing a right-handed Kaplan-type fermion for $r_{W,s} = 1$ and $m(s) = m \cdot \text{sign}(s)$. Below, we show the interval where $\lambda(p)$ takes its values for a given parameter M and $\alpha = \alpha_M$. (For $\alpha = -\alpha_{-M}$ the shaded area is mirrored at the origin.)